training and testing rears. This is admittedly a pretty small chunk of each S's behavior, but it will do for the case at hand.

Immediately noticeable in Table 19 is that most Ss showed a decrease in latency from training to test indicating that one hand-removal reinforcement shapes behavior. The other immediately apparent item is that the variable having whopper effects was cue change from training to test. In other words, the large reductions in latency came in the groups where stimulus conditions were not changed while the small gains in latency accrued to the change from a small number of cues to a large number or vice versa. It is also obvious that cues at the time of test *per se* had very little effect on behavior and cues in training only slightly more. The "real" effect is clearly the impact of change in cues or lack of it from training to test. In line with a huge number of generalization and generalization decrement studies, this experiment shows behavioral decrement associated with stimulus change.

The simplest analysis is, of course, a direct comparison of the behavior of all Ss treated with cue change with the responses of those having constant stimulation. There is overlap by only one case in the two distributions. So by any statistical token the two sets of behaviors are from different parent populations. For the present purpose,the factorial analysis needs doing. The results of this procedure are completely consistent with inspection in revealing very high significance for the interaction (change) between the training and test treatments. (Note Case C previously discussed.) Neither cues on training nor test show anything to speak of. As mentioned previously, this procedure seems like a lot of

work for the returns involved, but is probably worth it for didactic purposes.

Table 20 contains some data from a quite different experimental setting where the influence of partial reinforcement and success or failure in anagram solution was tried out. College students were first conditioned to emit a certain class of words with verbal reinforcement. Half were conditioned on a 100% and half on a 50% reinforcement schedule. The reinforcement groups were then further subdivided for anagram problems. Half of each subgroup was given insoluable anagrams with one letter changed so that a word could not be constructed (Failure). The other half had the solvable anagrams (Success). The final phase of the experiment consisted of conducting extinction for the originally conditioned word class. Thus the investigation was designed to test the effects of reinforcement schedule, success or failure and the joint action of the two. Table 20 contains a selected protion of the data chosen to represent the major trends of the original numbers.

A quick look at this table indicates a clear trend in the data. It is quite apparent that the experience of success or failure with the anagrams had little impact on behavior. It is also obvious that the joint action of the two variables had very little effect. The large effect is associated with reinforcement schedule. Only one case in the 100% group gets onto the 50% distribution indicating a far greater resistance to extinction after partial than 100% reinforcement. Either way the data are sliced - the simple Arrangement Technique or the complex factorial analysis - the outcome is significant for the reinforcement variable alone.

Table 20

Number of words emitted in extinction after 50% and 100% reinforcement in conditioning and after success or failure in solving anagrams.

REINFORCEMENT SCHEDULE	ANAGRAM SOLUTION	
	Success	Failure
	13	15
50%	16	12
	15	11
	9	8
100%	5	10
	12	4

5. Factorial Design: More Than Two Treatment Groups. Thus far we have considered only "simple" factorial designs where each of the two variables is broken down into only two conditions. From a design standpoint, of course, each dimension of variation may encompass three or more conditions. For example, in studying generalization of the size concept in children, one might train three groups, one on animals, one on "pure" shapes and one on toy vehicles. One third of each group would then be tested for generalization and generalization decrement on each of the three types of objects. The focus would be on the extent to which size discrimination with one set of objects generalized or transferred to the other objects. Or one might be interested in studying acquisition rate in children as a joint function of socio-economic status and age with several degrees of each variable represented.

In an experimental example, Rickard (1959) studied the extinction behavior of college students as a joint function of reinforcement schedule in conditioning and degree of cue change in extinction. The results are contained in Table 21 where it is apparent that both dimensions of variation had a large and consistent impact on behavior. It seems obvious that the big decremental effect from no cue change (UC) to the other extreme (EC) and from infrequent partial reinforcement to more frequent is so great that little interaction of the two dimensions could emerge. On more thorough analysis this turns out to be precisely the case. A large chunk of the variance is taken out by cue change and another large amount by reinforcement schedule leaving very little behavioral variation for interaction. Again inspection pays off. One might conclude, even with the

Table 21

Extinction responses in college students as a joint function of reinforcement schedule in conditioning and degree of cue change in extinction.

(Rickard, 1959)

SCHEDULE OF REINFORCEMENT	S	DEGREE OF CUE CHANGE UC	MC	EC
25%	1	68	32	37
	2	65	28	20
	3	61	26	11
	4	45	21	9
	5	40	16	9
	6	38	13	9
	7	28	11	9
	Mean	49.3	21.0	14.8
	Median	45.0	21.0	9.0
50%	1	69	21	8
	2	30	19	7
	3	24	6	6
	4	23	5	3
	5	12	2	3
	6	9	2	1
	7	5	0	0
	Mean	24.6	7.8	4.0
	Median	23.0	5.0	3.0
75%	1	35	36	19
	2	26	14	11
	3	24	8	3
	4	21	7	3
	5	13	3	2
	6	10	2	2
	7	3	1	2
	Mean	18.8	10.1	6.0
	Median	21.0	7.0	3.0

relatively large number of numbers represented in Table 21, that inspectional analysis could tell the whole story without any actual manipulation of the numbers. Such a procedure would save a great deal of time, effort and frustration on the part of the analyst.

To illustrate this type of design involving two dimensions of variation with three or more groups along one or both, an experiment dealing with the Freudian concept of "displacement" may be cited. Miller (1948) has presented an impressive translation of the concept of displacement into stimulus-response terms. His argument goes that the approach tendency exhibits greater generalization than the avoidance in an approach-avoidance conflict setting. From this position he deduced that experimentally induced conflict would be followed by an increment in response strength when the stimulus situation was changed. Miller and Kraeling (1952) found, in line with this expectation, that comparable groups of rats ran more frequently in changed than unchanged alleys after conflict training. A series of studies with pigeons in Skinner boxes failed to yield this incremental effect (Brush, et. al. 1952). More recently, Murray and Berkun (1955) attempted an integration of Miller's conflict (1951) and displacement models, and tested their deductions using training procedures very similar to those of Miller and Kraeling.

A re-examination of the Miller-Kraeling procedure seemed appropriate. Their rats were given approach training first, followed by avoidance conditioning. Thus, the terminal response learned was not-to-run. From the point of view of a contiguity theory, we can expect cue change to weaken the last response conditioned. If we consider the two mutually exclusive response classes of running and not-running, weakening of the

latter will produce an increment in the former. If this is the case, the Miller-Kraeling results can be accounted for by focusing on the effects of generalization decrement in the last response conditioned.

It follows from this line of argument that conflict is not as essential to the "displacement" phenomenon as it would seem to be in Miller's position. The effect should emerge under conditions of pure avoidance conditioning without the approach aspect. Lord and Taylor (1958) ran such a study in which groups of rats were trained in runways under conditions of 1) approach, 2) avoidance, and 3) combined approach-avoidance (conflict). Two runways were employed differing in height, width, interior brightness, dividing lines and texture of the floor and ceiling characteristics. The approach group is, of course, the traditional generalization and generalization decrement condition. Half of the straight approach group was trained to run for food in one alley; the other half in the other. Half of each of the sub-groups was tested in the same (training) alley and the other half switched for test. No alley differences emerged.

In the avoidance condition the rats were dropped into a padded bucket in the terminal unit of the runway. In the approach-avoidance condition, approach training was given first followed by the avoidance-drop treatment, following the Miller-Kraeling procedure.

Before turning to the outcome, it needs to be repeated that the key group is the straight avoidance group if the contiguity reasoning is correct that the terminal response is crucial and if this group exhibits increased running on test (changed-cue) trials, conflict is not an essential ingredient for the generation of the phenomenon.

The results of the Lord and Taylor "displacement" experiment are summarized, S-by-S, in Table 22 where focus on the "pure" avoidance group indicates a marked difference in behavior between the changed and unchanged conditions. In essence, the rats exposed to change, ran, while those remaining under their training stimulation did not. The contiguity argument proved out in the data and the case for conflict does not hold water. In passing, it is obvious that the straight generalization decrement approach group behaved exactly in accord with expectation.

The behavior of the conflict group is a focal point. In their case, running was first conditioned and then replaced by avoidant, non-running behavior. The expectation is,that under changed conditions, the last response trained (avoidance) should be weakened and replaced by the only other response (previously) conditioned, namely, running. In other words, under changed cue conditions the conflict rats should revert from avoidance to running. They did. The difference between behavior under the conflict-changed and avoidance-changed conditions is not great, but in the expected direction. Again the contiguity position is supported.

The argument might arise that the avoidance condition is actually an approach-avoidance conflict because the rat brings an "exploratory" approach tendency to the experimental setting that is pitted against the avoidance conditioning. If this situation prevails, Miller's position and ours reduce to the same thing. Our contiguity viewpoint still has the advantage, however, of reference to directly observable responses and stimulus changes rather than hypothetical entities. In any event, this is a tenuous argument. After all, any results can be explained in an ad hoc

Table 22

Median running time in sec. to reach the end box in three two-min. generalization test trials after approach, approach-avoidance (conflict) and avoidance conditioning.

(Lord and Taylor, 1958)

	APPROACH		CONFLICT		AVOIDANCE	
	UNCHANGED	CHANGED	UNCHANGED	CHANGED	UNCHANGED	CHANGED
	5	4	360+	360+	322	118
	2	50	264	290	360+	110
	1	19	360+	67	360+	360+
	2	32	360+	53	360+	135
	5	8	360+	25		
		12				
Median	2.0	15.5	360+	67	360+	126.5
P(t)	.024		.016		.016	

fashion post facto, but this procedure is not likely to advance behavioral science.

Turning to the numbers of Table 22, this is a case where the interaction of the two dimensions of variation is of no great consequence. We are interested directly in decremental effects in running in the "pure" approach condition and in decrements in not-running (avoidance) in the conflict and "pure" avoidance conditions. By this token, the pertinent, direct comparisons are made within and across groups. The main findings are clear without statistical manipulation: running decreases in the approach group and increases in the other two groups. The results support the contiguity position and, minimally, raise serious doubts about the need for conflict in the occurence of the displacement phenomenon. In turn, of course, the theory behind the conflict position is called into grave question.

Given the theoretical issues involved and the data of Table 22, it is left to the reader to decide whether traditional, interactional statistics should be applied to the data.

Another illustration of "simple" factorial design may be found in Newton's (1953) investigation of the effects of reward and punishment on learning and tachistoscopic recognition. He somewhat unintentionally conducted a test of the Skaggs-Robinson Hypothesis previously treated. He had three groups of 20 college students each learn a list of five-letter meaningful words. One group was presented verbal and monetary rewards for correct responses; a second was verbally chastised and lost money for incorrect responses; and the third group constituted a baseline case receiv-

ing neither reward nor punishment. Next he ran a tachistoscopic test in which he briefly flashed the original words singly along with words having one, two, three or four letters changed. For example, if the original word happened to be BASIN, it was presented along with BASIS, BARON, and BIRCH. On a generalization basis, it can be expected that the more similar the word to the originally learned one, the greater the probability of the original response. Thus with one or two letters changed, intrusive errors of this kind should be maximal. With maximal dissimilarity, on the other hand, discrimination should operate to appreciably reduce interference effects. There is practically no incompatibility between BASIN and ALONE presented tachistoscopically. The original word should, of course, benefit from previous practice in the acquisition setting. From these premises, it follows (post facto) that a Skaggs-Robinson function should emerge as dissimilarity of stimulus materials increases. The following insert shows the mean number of errors in the tachistoscopic test as a joint function of learning condition and number of letters changed:

LEARNING CONDITION	NUMBER OF LETTERS CHANGED			
	0	1	2	3-4
Reward	2.2	4.9	4.3	3.3
Ignore	1.7	4.2	4.2	3.4
Punish	4.0	8.8	8.3	5.4

The Skaggs-Robinson Hypothesis is supported by the emergence of the predicted, asymmetrical U-shaped function. Errors are minimal for the originally learned words, next for the most dissimilar words and most

frequent for the intermediate degrees of change. NOteworthy, but incidental, is the fact that the tachistoscopic presentation was sensitive to the punishment treatment where the original learning situation had not been.

By token of the several previous presentations, it should be obvious that, in this instance, all three dimensions of variation turned out to be significant: learning condition of reward-punish-ignore, number of letters changed in tachistoscopic recognition and the interaction of the two main treatments.

MAGNITUDE: MORE THAN TWO "SIMULTANEOUS" DIMENSIONS OF EXPERIMENTAL VARIATION

Sometimes investigators, possibly unfortunately, decide to throw a number of experimental vegetables into the design stew at the same time. There are alternate strategies. One is to piece out the research with a number of sub-experiments with their cross-comparisons and their economical shortcuts of pivoting several experimental groups on one control. This is clearly the present position. The alternative is to throw all variables into the pot at the same time and see what emerges. A case supporting the latter view can clearly be made, but there are a couple of objections. One is statistical, namely, that anova procedures are quite (maybe overly) sensitive to outlying cases and other data aberrations so that the analysis can be thrown off and distorted, leading to inappropriate conclusions. The other objection - in addition to complications of design and statistics - is more behavioral, namely, that with many variables

applied together it is frequently difficult to "see" what has happened. The human eye obviously has its limitations and it is very difficult if not impossible to determine what differential action has occurred when, say four or five variables and their several interactions are operating.

With these comments made, let us consider the kinds of experimental cases in which complex anova might be applied where three or more dimensions of experimental variation are employed. In an educational setting one might be interested in studying learning rate as it covaries with age, sex, socio-economic status, sibling position, presence or absence of parent(s), and characteristics of the examiner. This project could be accomplished by a series of experiments based on comparable samples of Ss involving one or two variables at a time and cross comparisons along the relevant dimensions or all variables could be applied together. The latter involves considerable pre-experimental planning. With the six variables cited there are a minimum of 15 sub-groups to consider. The problem of procurement is real and becomes more pressing when it becomes apparent that variability in an experiment such as this is likely to be great and fairly large Ns needed.

Another case in point consists of some data that came to hand recently involving the covariation of grade, sex, birth order and their interactions against IQ, achievement and scores in reading, language, arithmetic and total score along with Grade Point Average. Out of this veritable hodge-podge (some measures were missing for some Ss) there emerge some 42 F-values from anova along with various and sundry other numbers. Some of the values are significant, some insignificant. Replication yielded

comparable findings. In fairness to the investigator, it should be noted that the variables and relationships he focused on held up across the two studies. The simpler way would clearly have been to select out a couple of variables and a couple of measures and run the partial (but main) incomplete block design utilizing only a small portion of the variables.

It would seem that, unless the investigator really doesn't "know" the effects of his several variables and is therefore simply indulging in experimental fishing, it is wise to put the effort into judicious selection of variables and measures and cut the design down to workable size.

To illustrate complex anova, let us consider a hypothetical case.

Sex and The Squirm Test. Suppose one were interested in the influence of sex content in a motion picture on behavior. Obvious variables to build into the design are the sex and age of the audience. One way to conduct the experiment is with large Ns; another with small. We will take the latter. Individuals or groups might be matched on the basis of previous exposure to movies with and without sex content and the like, but for simplicity purposes we will take independent groups with sub-Ns of five. There emerges a 2 X 2 X 2 factorial design with two conditions of each of three dimensions of variation: 1) movie with and without sex content (one would take appropriate control action with regard to duration, other content, seating arrangements, etc.), 2) Male and female participants and 3) older and younger Ss. Thus there would be eight groups of five Ss each.

A major consideration, as always, is the index of behavior to be employed. Attitudes toward the movie are one facet of behavior, but a

more direct index is called for. The measure proposed is the Squirming Response consisting of the amount of movement exhibited by the audience in their seats. It would be relatively easy to doctor the individual seats with recording devices that would be sensitive to and pick up slight body movements. If one wished to get fancy about recording, photographic records could be taken of the facial expressions and movements of each individual S, but the short-coming of this procedure, as always in this instance, is the enormous effort that has to be expended in analyzing the data. Records of squirming would include body contacts with neighbors, particularly of the opposite sex and these possibly should be partialled out for separate treatment, but a line has to be drawn somewhere.

Hypothetical data are presented in Table 23 where careful inspection shows more activity for the movie with sex content and in that framework more movement for males and younger people. There is thus a clear suggestion of an interaction effect between the movie variable and audience sex and age.

Sub-comparisons by the several techniques previously spelled out are clearly appropriate, but for purposes of exposition the writer waded through the classical complex anova procedure for treating these data. (Incidentally, it took a long half hour to complete the analysis plus some 45 minutes of a graduate assistant's time in replicating the analysis. All sub-comparisons by more efficient shortcut techniques took a total of less than 15 minutes; looking at column and row sums for inferences took less than five minutes.) From the overall analysis for main effects only the movie emerged significant. The major variables of sex and age of audience

Table 23

SEX AND THE SQUIRMING RESPONSE

Hypothetical squirming responses to a movie with and without a heavy sex element for audiences split by age and sex.

MOVIE WITH SEX				MOVIE WITHOUT SEX			
Male		Female		Male		Female	
Young	Old	Young	Old	Young	Old	Young	Old
9	7	8	4	3	6	4	7
8	5	8	4	2	5	4	6
7	4	7	3	2	5	3	6
7	4	6	3	2	4	2	5
7	3	5	1	1	3	2	4

did not hold up. This finding makes sense in terms of the raw data of Table 23. Here it is clear that the totals for sex alone and for age alone are approximately equal indicating practically no effect of these variables.

On the interaction side of the fence it was, however, a different story. In line with inspection the first order interactions of movie with audience sex and movie with audience age both turned up to be highly significant. The age-movie interaction had considerably more impact and accounted for much more of the variance than did the movie-sex interaction in conformity with expectation from inspectional analysis. The second order interaction of all three variables was quite insignificant.

This experimental example of sex and the squirm test is a relatively simple one and the data have been doctored to yield clearcut effects. Such is not the case with many other instances where single deviant cases or peculiarities of interaction turn up. This point will now be illustrated.

It is quite an easy matter to criticise published articles - even one's own - on methodological and statistical grounds. As a matter of fact it is relatively easy (and dramatic for undergarduates) to open their textbook to any page and find something wrong. The words are easy; the numbers are hard. For many years I have given students in graduate research and methodology seminars articles in Japanese with the purpose of reconstructing the experiment from the tables of data. A moderately sophisticated observer of the behavioral scene can do this with little trouble. Another tour de force I have conducted is to have students open any issue

of any journal to any page containing tabular material and work it over. In roughly two-thirds of the cases we have been able to find something "wrong", mostly minor, but sometimes major.

I found a copy of the Journal of Experimental Child Psychology for October, 1966. It opened most easily to page 253 which revealed a table, Table 2, entitled "Mean number of responses made during each minute of reinforcement period". I examined the data without referring to other parts of the article and drew my conclusions without manipulating the averages. (No indices of variability were presented). I then looked at the authors' (Stevenson and Odom) conclusions. They did not agree with my interpretations so I went back and looked at the "problem" and "method" sections of the paper. Some quite interesting angles emerged. It turned out to be a rather complex experiment entitled "Visual reinforcement with children". In it the investigators tried out the effects of Examiner sex differences along with the sex and age of Ss who were 192 boys and girls ages 6-7 or 10-11. The task to be learned was a lever-pressing one with pictures, colors and line drawings presented as reinforcement on the average of once every 20 responses. There were thus two conditions for each of three dimensions of variation and three conditions for the reinforcement dimension.

Initially, in examining these data my focus was on the treatment variables, but it shifted after consideration of their data concerning operant level performance ("base rates"). Table 24 contains the mean base rates and, parenthetically, the comparable figures for the period of reinforcement.

Table 24

"Mean base rates" (operant level) of lever-pressing in 192 children as a function of age of S and sex of E and S. The figures in parentheses are corresponding means for the reinforcement period when pictures were presented.

(Stevenson and Odom, 1964)

	SEX OF E	
SEX OF S	Male	Female
Male		
6-7 years	72.0 (72.3)	59.2 (69.1)
10-11 years	78.2 (82.9)	56.3 (70.4)
Female		
6-7 years	63.0 (60.1)	55.2 (64.2)
10-11 years	71.9 (73.3)	46.2 (50.0)

The eye-catching feature of the numbers in Table 24 are their similarity. In other words, if learning took place, the indications for it are limited. The average gain in means from operant level determinations to conditioning was ca. five responses, a matter of a shade over 7%. One cannot fail to be impressed with the small order of magnitude of these numbers. By no stretch of the imagination do they indicate an appreciable amount of acquisition. To complete the picture it would be nice to have at hand the count figure of the number of Ss showing increments in responding during reinforcement. Another special feature of these findings is that in extinction a number of the groups showed an increase in reaction contrary to the usual decremental effects.

The other noteworthy item in Table 24 is the Examiner variable. Far and away the largest differences contained in this representation are associated with this source of variation. This point holds across age and sex levels of Ss and across from operant level through conditioning into extinction. Response level was higher with the male E than with the female. The investigators' elaborate analyses of the data support what the naked eye can see clearly. The authors report no other significant differences in the data although inspection shows a consistent trend for male groups to respond at a higher level than female. All eight of the differences for comparable means in Table 24 are in this same direction.

The point clearly to be made in connection with this paper is not that the investigation was "wrong" or valueless or anything of the sort. It is clearly a worthwhile piece of research. It seems apparent that the data were not exhaustively examined and the small differences between un-

conditioned and conditioned responding were not taken into account. This matter does not effect the conclusion concerning Examiner effects directly. Indirectly, it makes one wonder about the potential generality of the Examiner differences. It is important to note the small amount of learning accruing to these procedures. It raises a number of solid parametric questions concerning the basis for minimal acquisition.

The data of this study point up and capitalize the basic need for careful study of numbers reflecting behavioral changes before elegant statistics are applied. A major facet of the analysis of this study should have focused on the pre-conditioning-conditioning differences. Inspection would have pointed the way.

It is frequently a profitable exercise to lay out a research program on a grand scale throwing all the variables and potential variables into the pot - on paper. At this point it is also a worthwhile intellectual and didactic exercise to spell out the large scale design that would be translated into experimental practice given adequate funds, time and personnel. Having laid out this overview, it is extremely wise to then study it carefully to insure that the variables included are worth the experimental trouble, that others haven't investigated them thoroughly, and so forth. What frequently happens is that half or more of the grand design can be sloughed off ab initio. Further consideration of the remaining matrix of variables sometimes suggests a priority listing such that some subexperiments emerge as more basic and appealing to the investigator than others. Wading through this somewhat tortuous process may leave only 10% of the original overall program, but it will be the heart of the matter

for the particular investigator at that particular point in his research career.

Again, caveat emptor re large numbers. Large numbers of variables have many of the same disadvantages as large numbers of Ss plus the fact that the situation is sensitized to enhance the role of chance by their use. Quantity by no means insures quality. As a matter of fact it may well mitigate against it.

MAGNITUDE: ANALYSIS OF COVARIANCE (ANCOVA): PARTIALLING OUT THE EFFECTS OF ONE VARIABLE UPON ANOTHER

There are many instances in behavioral research where a variable influences behavior that is not part and parcel of the experimental treatment. The investigator is sometimes "aware" of the action of such variables and in such instances, of course, makes his behavioral measurements. There are basically two cases of this kind. The first is where the investigator has introduced a pre-test or selection or matching variable and for unknown reasons (presumably "chance"), the situation goes awry and the groups do not come out equivalent on the initial measure. When this lack of equivalence is large, and particularly when it stacks the deck in favor of the experimenter's hypothesis and expected direction of his treatment, some corrective procedure has to be applied. Since these events are post facto, i.e., occur after the initial measurements, the corrective is statistical. Verbally the technique is straightforward. The initial and final measures are subjected to independent statistical analysis. Then the relationship (correlation) between the two is statistically handled in

such a way as to partial out differences in the initial measures from those in the final behaviors. The arithmetic is a little complicated, but this is the essence of the procedure. For example, a substantial difference in terminal behavior may be washed out when an appreciable initial difference in the same direction is taken into account.

The second instance in which initial differences have to be considered in assessing final ones involves the case where the nature of the experimental treatment is such that it has an appreciable impact on both initial and final measures. (A related case is that of a variable that free-floats and is not under experimental control where E can simply measure its behavioral consequences). For instance, certain treatments, e.g., partial reinforcement and distribution of practice have notable influence on both conditioning and extinction. The investigator may be interested in the "pure" effect of such treatments on extinction, say, uncontaminated by the influence of the procedure on conditioning. In this instance, he may wish to remove the effects, statistically, of the influence of the variable on the earlier measures from that on the latter.

There is one special but basic problem that must be considered in this context. It makes a great deal of difference in which direction the experimental treatment influences the two phases of measurement. It may increase response strength in both, decrease it in both or act differentially to increase it in the one and decrease it in the other or vice versa. A clear case in point is partial reinforcement. Here lowered frequency of reinforcement in conditioning tends to retard learning and generate a lower level of stabilized responding after conditioning is complete as contrasted with a higher frequency of occurrence of the reinforcing stimulus.

In extinction, however, the situation is exactly reversed and the group with the lower frequency of reinforcement yields greater resistance to extinction than the comparison group with more reinforcement in conditioning. If one is focusing on extinction there is no problem. Differences favoring greater resistance to extinction in the partially reinforced group emerge despite this group's lower level of performance in extinction. If one is a purist, one might wish to still apply the analysis of covariance procedures, but by the nature of the situation such application can do nothing but merely enhance the extinction differences favoring the partially reinforced group. In the limiting case where non-overlapping distributions appear, correction is clearly a waste of time.

Some previously unpublished data illustrate this point. The reasoning behind this experiment was that, on a generalization and generalization decrement basis, the more conditioning is made like extinction, the greater the resistance to extinction. The two essential ingredients of prolonged extinction are absence of the reinforcing stimulus and a low level of responding. These were approximated in conditioning by teaching pigeons to wait between responses. (This could be described as an experiment in damping out "impulsivity".) This training was not easy because of the ballistic nature of the pecking response. In training, reinforcement was presented for the E-group only after longer and longer pauses. This shaping was continued until the E-birds were making about 30 responses per hour or one every two minutes on a partial reinforcement basis.

The Control Group is a problem in an experiment such as this. To use the typical 100% reinforcement control seems like working in an entire-

ly different universe. In conditioning the control group would be pecking and eating most of the time while the E-group would be standing around waiting. The situation could be compared to a baseball player and violinist in that they both use their hands, but in obviously different capacities. For this reason it was decided to use a fairly infrequent aperiodically reinforced group as the control so that the two sets of data in extinction could be plotted on the same axes.

The results are summarized in Table 25 where it can be seen that there is little comparability between the E- and C-groups in either conditioning or extinction. Conditioning responses for the Control birds exceed those of the E-birds by a factor of 60 and in extinction by one of nearly 10. There are two different ways of tackling these data. The hard way is to perform the ancova testing for significance in conditioning and in extinction separately and then teasing out the influence of the first on the second by way of the correlation between the two sets of data. This involves a lot of arithmetic. The easy way was the one followed, namely, to percentagize each bird's extinction behavior over his conditioning behavior. Actually, in this instance ratios were simply taken of per-hour performance in extinction over the same figure for conditioning. This procedure accomplished at least two things: it cuts back on variability and it takes into account any correlation extent between conditioning and extinction. In a sense it is a simple form of ancova.

The results are dramatically reversed. In the ratio figures the E-birds exceed the C ones by a factor of 10. It might be noted in passing that the treatment "worked" in the sense that the birds for which condi-

Table 25

Percentagizing behavior: Conditioning and extinction responses per hour for a C- and E-group where conditioning was made like extinction for the E-group.

(Jenkins, 1955)

CONDITIONING		EXTINCTION		EXTINCTION/CONDITIONING	
C	E	C	E	C	E
3500	23	154	10	4.4	43.5
2700	29	115	12	4.3	41.4
2200	37	83	17	3.8	45.9

tioning was made like extinction continued to respond and even after 36 hours of extinction were clipping along at about one-third of their conditioning rate - and well above operant level.

Analysis of the data of Table 25 is quite straightforward. The Arrangement Technique yields a P-value of .05 for the two non-overlapping sets of three events in the ratio figures. If one wishes the arithmetical exercise, the classical t-value is greater than 25 and is quite clearly highly significant for the four degrees of freedom involved. The Range Test is definitely appropriate to these ratios and yields a highly significant value exceeding 40.

In this instance there is serious question whether the classical ancova is applicable. A correlational term based on two sets of three cases had very little relational meaning. It is recommended that wherever the data resemble those of Table 25, some form of percentagizing procedure be employed both for simplicity's sake and for that of statistical sensitivity and minimal arithmetical error.

Basic questions have previously been raised about complex designs chiefly concerning the interpretation of the data of such items as higher-order interactions. This same criticism and caution applies to the ancova situation. For instance, significance may not emerge in the "before" measures, but the behavior may head in the same direction here as in the after measures so as to spuriously enhance significance in the "after" effects. Again, correlations can be slippery things and the ancova case is no exception. For example, suppose highly differential correlation exists between the experimental and control groups across the "before" and "after"

measures. Should one combine them anyway disregarding the differential or is it reasonable to convert to z-scores and combine? Suppose the correlation for the C-condition is positive and that for E negative and the difference in correlations is significant by usual standards? These and many other questions should make one think several times, not just before applying ancova, but more importantly before designing an experiment appropriate to the ancova procedure.

As a case in point during World War II an investigator, quite logically, tried out a new pilot selection instrument by testing neophytes and skilled pilots on it. The two distributions of measurements practically did not overlap. When the two sets of scores were combined and the overall distribution correlated with an outside criterion, the resulting correlation was high and positive although it was near zero for each group separately. By combining two distributions apparently drawn from quite different parent populations, a markedly spurious correlation was generated, as witnessed by the low correlations produced by taking each group individually against the outside criterion. This situation illustrates the kind of complication ancova can run into. More pertinent examples will be presented later.

To return to the ancova type set-up reference back to Table 17 is another case in point. It exemplifies a repeated measurements arrangement in which one group of birds was trained and tested under distributed practice conditions and the other under massed. Ancova could be applied in one of two ways to these data. First, only the behavior of the first extinction session could be partialled out of that of the third extinction

session so that initial differences favoring the massed condition would be corrected for in the reactions of the third extinction session also favoring the massed condition. (The percentagizing procedure from first to third sessions was suggested in that context and it will be recalled that differences indicating faster extinction for the distributed condition held up after conversion to percentages.) The other way in which ancova could be applied to the distribution of extinction experiment constitutes the most complex analysis that can be applied to these data. It involves partialling out extinction behavior in the first two sessions from that in the third session. This step becomes quickly tricky and sticky because it involves appreciably increased variability and differential correlation not only across the distribution of extinction variable, but also from the 1-3 and 2-3 sessions correlations. Application of ancova is likely in this instance to result in a mishmash, statistically speaking. Percentagizing seems like far and away the most efficient and statistically sensitive procedure for the numbers of Table 17.

The overall point here is that any repeated measurement set-up where initial differences emerge can be considered an ancova arrangement. On a few rare occasions, traditional ancova may be necessary, but in most instances a percentage conversion procedure will do the job faster and more efficiently. It also follows that any classical transfer of training design involving before and after-treatment measures can turn into an ancova set-up if the investigator gets a bad break and the groups do not turn out to be initially equivalent. Savings scores were designed to take care of this complication and do.

To illustrate the complications of the ancova design, a hypothetical (but not unrealistic) experiment was designed involving the application of four different science curricula to second-grade public school pupils. Bypassing the large difficulties involved in selection and constructing the curricula, training the teachers and various other pieces of administrative spade work, it is assumed that a well-designed study was laid out in which pre-experimental science knowledge was determined by a pre-treatment assessment measure and an IQ index of intellectual ability was employed. After exposure to the experimental curricula, a post-treatment measure was applied to reflect changes in science knowledge, methodology, attitude and philosophy.

While it clearly would be far more reflective of the data to have individual scores on at least a sub-sample, we will settle for means and standard deviations as representative of trends in the hypothetical data. This information is summarized in Table 26 as it would be summarized in a journal article.

The data of Table 26 fall within very realistic limits for this type of experiment. It might be noted in passing that the hypothetical results have not been complicated by differences that well might emerge such as socio-economic status, number in family, residence and presence or absence or parents. The first noteworthy item is the considerably greater loss of Ss in the Gamma group as contrasted with the others. Since all groups started with an N of 100 this attrition contributes an appreciable unknown and unfortunate bias to the data. They presumably would be analysed anyway.

Table 26

Hypothetical data on the influence of four different science curricula on a science post-curriculum test in the second grade. A pre-curriculum test and an IQ test were given initially.

	SCIENCE CURRICULUM			
	Alpha	Beta	Gamma	Omega
Initial N	100	100	100	100
Final N	95	98	63	94
Pre-Test Mean	50.3	49.8	60.7	54.2
SD	12.5	15.9	20.5	17.8
IQ Mean	98.4	101.3	108.7	103.6
SD	14.8	18.7	13.4	15.3
Post-Test Mean	51.7	56.9	73.4	75.7
SD	10.4	14.6	18.9	22.3

Turning to the main findings, the post-treatment test performance, there seems to be an appreciable difference in the extremes (Alpha vs. Omega) of the rough order of one and one-half standard deviation units. This in itself should be significant, but there are a number of other considerations before any conclusions can be drawn. The high wastage rate in Gamma cannot be ignored, but the alternative is to throw the whole group out and this seems frightfully wasteful, particularly since experiments of this kind take at least a year to plan and another year to conduct.

Of considerable import are the relatively large differences apparent in the mean science pre-treatment scores. The maximal difference amounts to about half a standard deviation, a magnitude not to be ignored. More basically, Gamma and Omega, which have the highest pre-treatment test scores, also have the highest post-treatment scores. A correction must be introduced for this event since post-treatment differences may reflect in large part differences in initial knowledge of the pupils regarding science. Furthermore, mean IQ's show the same trend, that is, the groups which have the higher post-treatment test scores also have the higher initial IQ averages. Again some correction must be introduced or the final differences may simply reflect greater intellectual ability alone or combined with greater science information.

The data are incomplete for an analysis of covariance as they stand. The correlation (or at least the S-by-S data or the cross products) are needed between pretest and IQ on the one hand and post-treatment performance on the other. Given incomplete information, let us see what can be retrieved.

While the post-test scores differ appreciably across groups, it seems likely that they are contaminated by pre-test and IQ differences. Considering two sets of differences simultaneously - across groups or conditions and across tests or measures - it is obvious that, although the post-test differences are of the order of 1.5 sigma units, the pre-test and IQ differences are of the order of around a half a sigma. Furthermore, it is quite reasonable to assume a fairly substantial positive correlation between scores on the two initial devices and the post-treatment index. Given this information and these assumptions, it seems quite plausible to expect that the post-treatment differences will be highly diluted when pre-treatment test scores and IQ are partialled out. From a practical standpoint, one must - in the absence of other considerations - recommend treatment Omega for use, since the greatest science test gains accrued to it, the sample stayed fairly intact and the average IQ is not too far out of line with the lower groups. Gamma has the disadvantages of high attrition for unknown reasons, the highest mean IQ and an appreciably lesser gain on the science measure.

Replication is clearly called for and the refinements are obvious. Smaller Ns should be employed with groups matched on pre-treatment test scores and IQ (if these relate to post-treatment performance on an appreciable scale), and some consideration of socio-economic features and associated items along with factors contributing to attrition. If Alpha and Beta correlate substantially, one should be dropped and the one retained that takes the lesser time and effort to teach the teachers and to administer. With this matched group design, direct comparison could

be made of post-treatment test performance.

It is time to pull together the essential steps in analysis of covariance. First, it should be noted that the ancova procedure adjusts the effects of the experimental treatment, that is, the post-treatment measure of behavior, according to behavior on the pre-test or pre-treatment measures. The extent of adjustment depends on three items: 1) the extent of correlation between the pre- and post-treatment measures; 2) the size of the difference in the groups on the pre-treatment measure; and 3) the magnitude of the difference in behavior on the post-treatment index.

The actual steps in ancova, while somewhat cumbersome and lengthy arithmetically, are quite straightforward. First analysis of variance is accomplished on the post-treatment measures; then on the pre-treatment ones. Next anova is applied to the cross-products of the two measures. Finally, in the ancova analysis, the post-treatment differences (on which the experimental focus falls) are adjusted or corrected for the pre-treatment differences and the correlation between pre and post measures.

A couple of side comments are called for. Ancova assumes linear regression, i.e., a linear relationship between the variables involved. This is sometimes the case, sometimes not. When non-linear regression prevails, real problems are posed for ancova. A second point concerns the use of more than one supplementary or pre-treatment measure as in the educational experiment depicted in Table 26. Matching or selection on more than one pre-treatment variable rapidly reaches a point of diminishing experimental returns. As a matter of fact it is rare that more than one variable will appreciably contribute to the picture and the use of more than one many

times complicates the situation not only arithmetically, but compounds and confounds the complications of relationships among the several measures. More than one matching or selection variable is not recommended unless the very special case arises of two relatively uncorrelated indices both of which correlate with the criterion or post-treatment measure. If more than one variable is necessary, matching by pairs or by groups beforehand is far more efficient than ancova.

At this juncture we are ready for a real-life experimental example that demonstrates the complication, difficulties and short-comings of ancova as well as pointing up matters of design in the ancova setting. W.E. Morris (1953) experimentally examined the problem of teaching the analysis of language or communication of written messages by way of Charles Morris' types of discourse. He employed the traditional transfer of training model with a pre-test, treatment and post-test. He first pre-tested university graduate and undergraduate students on comprehension of written discourse. The experimental groups were then given training on Charles Morris' types of discourse while the control groups were exposed to the materials of Feigl and Osgood. (It should be noted that it was considered unnecessary to run the "pure" control groups of (no treatment whatsoever) and thus the cards ab initio were stacked for reduced differences across groups.) A post-treatment test of "language comprehension" was then administered with time devoted to the post-test recorded. Post facto no systematic differences emerged among the various control conditions and they were lumped to simplify the analysis. Three experiments were conducted; the first involved 10 Morris Ss and 18 controls ("Others"); the

second 6 and 17; and the third 9 and 9.

In this ancova design there are two predictor, selector, matching or co-variables: pre-test or pre-treatment performance and time on post-test. (It is obvious that all other things equal an S devoting two minutes to the complex post-treatment task is not likely to do as well as one giving it 30 minutes. The experimenter found it took him 30 min. to complete the post-test.) The criterion is, of course, post-treatment test performance. The basic experimental issue is whether or not the Morris treatment contributed more to the understanding of language than did the other procedures.

There would not be problems of treating the data from these experiments if pre-treatment performance and time on the post-test were equivalent acorss the Experimental and Control groups. In this case a direct and simple statistical comparison could be made on post-test performance. But they were not comparable, and recourse had to be made to ancova. (A couple of relatively minor complications will be ignored such as loss of a S or two who do not comprehend the instructions and some heterogeneity of variance.)

Table 27 summarizes the results from the three separate experiments. An overview of this vast mass of numbers suggests little systematic trend in the data. (The clear superiority of Ss i Experiment III can be ignored for the present purposes since it is attributable to refinements in technique and the use of graduate students.) More thorough examination of the data suggests higher post-treatment performance in the Morris group although the effect is by no means glaring. The statistical problem is to pin this apparent difference down in the light of other covariations. Or at

Table 27

An empirical example of ancova: W.E. Morris experiments on Charles Morris' types of discourse (1953).

	MORRIS TREATMENT	FEIGL, OSGOOD AND OTHER TREATMENT
EXPERIMENT I		
N	10	18
Mean Post-Test	19.7	16.8
Mean Pre-Test	15.3	15.2
Mean Time	15.3	13.9
EXPERIMENT II		
N	6	17
Mean Post-Test	19.0	12.1
Mean Pre-Test	17.0	10.5
Mean Time	13.5	13.5
EXPERIMENT III		
N	9	9
Mean Post-Test	31.6	31.9
Mean Pre-Test	29.6	28.4
Mean Time	18.4	11.2

least the data must be squeezed dry to see what, if anything, happened to the behavior of the Ss differentially treated by the different language systems.

The first complication arises because the differences in post-test time, though small, are apparently real in a statistical sense. Thus they must be taken into account in the final analysis.

A second and somewhat more basic difficulty arises when the correlations among the several measures are examined. W.E. Morris correlated pre-test performance and post-test time with the criterion of post-treatment performance. While the Ns are relatively small, the trends are clear. As a sample, the intercorrelations for Experiment I follow.

	MORRIS			OTHERS	
	Pre-test	Time		Pre-test	Time
Post-test	.68	.08	Post-test	.13	-.08
Pre-test	--	-.05	Pre-test	--	-.53

A clearer set of differential correlations is hard to come by. In the Morris group pre- and post-test correlated high positive; in Others it was near zero. The pre-test-time correlation in the Morris condition was near zero while it was substantially negative in the control case. Serious doubts are immediately raised about the legitimacy of combining the two sets of correlations for analysis purposes.

For each experiment, W.E. Morris separately computed the multiple correlation between the predictors and criterion, that is between pre-treatment test performance and time on the one hand and post-treatment

test scores on the other. The multiples were calculated, of course, separately for the Morris and Other conditions. These multiple correlations follow:

EXPERIMENT	MORRIS TREATMENT	OTHER TREATMENT
I	.68	.13
II	.84	.10
III	.73	.385

Again, the contrast in these relationships between the two conditions is striking. The Morris' multiples are quite high; the Control figures relatively low. The finding that these two sets of data are clearly drawn from different parent populations, dictates the use of some statistical technique other than ancova. But as in many complex experiments, the investigator wishes to test out his data from every angle so ancova was applied by way of correcting or adjusting post-treatment test scores for both pre-treatment performance and time devoted to the post-test. None of the overall outcomes exceeded the 20% level of significance.

At this point investigators are faced with several possibilities. The obvious solution is to design and conduct an improved experiment. (One ridiculous possibility is to forget the whole thing - a hasty mistake that might be made if follow-up analyses were not conducted.) There are however, other considerations that dictate communication of the state of the art to the public or, at least to that portion of it known as the graduate school. To exhaustively analyze the data after application of these complex procedures, Morris selected out two sets of eight Ss each who

matched up on pre-treatment test performance and on time employed in the post-treatment test. One set of eight was from the Morris condition and the other from the Control. The two samples turned out to be nearly perfectly matched on these two dimensions and a difference of nearly two to one emerged in post-treatment test performance: 26.8 versus 13.8. The two distributions of test scores overlapped very little and the difference by a t-test was significant at about the 1% level. One might read this as salvaging data - 16 Ss retrieved from an initial total of 69 - but it appears to be far more than that. It demonstrates that in the complex area of language interpretation where few behavioral principles are known, one system of analyzing communication is superior to others. At the very least it sets the stage for a program of research in this area.

W.E. Morris introduced one "experimental" gimmick that did not appear in the dissertation. After wading through the seemingly innumerable and complex analyses, he went back and culled out Ss he knew personally. These he sorted into two piles: those he judged to "like" him and those he judged not to. Comparing the performance of these two groups yielded some intriguing data. Not only did the "not-likes" take appreciably less time in performing on the post-test, but they scored at a lower level across the board. Following up on this finding (?) an eye-ball examination of the data indicated that almost all experimental Ss who spend 5 min. or more per item on the post-test performed significantly above chance while those spending less than 5 min. did not. On the basis of this reasonable if intriguing finding, an additional experiment was carried out. Using the two most reliable test items of six (decreasing the time required for testing),

Morris cajoled (?) the students of an undergraduate psychology class into spending a reasonable amount of time at the task. The results of this experiment clearly favored the Morris (experimental) group.

Attribute those findings to a "personality factor" or whatever, there seems to be a basic dimension here for experimental examination - and not just with human Ss. Rats do not seem to "like" certain investigators as witness increased frequency of biting, struggling, escaping from E, and very likely special behaviors in the experimental setting. These special behavioral features probably are more rampant with more complex species such as chimpanzees and college students.

AFTERTHOUGHTS, ODDS AND ENDS AND SOME OVERVIEW MATTERS IN EXPERIMENTAL DESIGN, METHODOLOGY AND STATISTICS

It is not always easy to buck the tide. Many writers today claim that improvements and refinements in statistical procedures put investigators in a position to lay out more elaborately designed experiments that yield vastly increased information. This is a moot point with which the writer disagrees wholeheartedly. The growth of a science is reflected at first in its increasing complexity of methodology and theory. Where one doesn't know much, one speculates widely and tries out innumerable things. As science progresses, relative simplicity sets in. The word "relative" is used advisedly. Einstein's basic formulation was superficially simple, but enormously complicated in its implications and ramifications. This "principle of relative simplicity" holds for both design and analytical procedures. Physical scientists are dealing with whopper

effects; they don't need statistical manipulations to tease the phenomena out. Their order of magnitude is 10 to 1 or 100 to 1 or a million to one.

This point clearly applies to the behavioral sciences. Right now there's a lot we don't know - although it would seem that we know a good deal more than we sometimes profess. We do have basic principles for manipulating and changing behavior on a large scale - although we don't always use them, particularly in the practices of child training and education. Much of our research, viz. psychotherapy, is a fishing expedition. We're trying to find variables that change individual behavior and measures of behavior that reflect our experimental applications. At the same time we're trying to construct theoretical structures that will handle the rapidly accumulating data.

There are two ways - at different ends of the continuum but not basically opposed - for tackling these problems. One is the overview approach in which any variable remotely related to behavioral change is thrown into the experimental pot along with variables that have theoretical or empirical foundation. This approach clearly involves complex design and elaborate statistics for partialling out the effects and in er-effects of the several treatments. The alternative is the classical method of univariable experimentation where one treatment is applied at a time. This procedure is traditional in the fields of sensation, perception, verbal learning a la McGeoch, comparative psychology and a few other areas. (The "pure" psychophysicist trains and uses one S and sometimes one other for replication.) The multi-variable approach seems more characteristic of some of the newer disciplines such as clinical psychology and the social

and educational fields. The nature of the problem at hand and the state of knowledge behind it will, of course, play a considerable role in determining which approach will be employed. The current view is that when we don't know much, we should play it simple and where we know a good deal we are in a position to do the same. After all very few (if any) of the great discoveries of science came about by way of multi-variable experimentation. It can be argued, of course, that this is an historical and cultural artifact. There is no answer to this; history can't be experimented on.

There are many pressing major problem areas in behavioral science among which may be mentioned "mental illness" and psychotherapy, child rearing and educational practices, that nebulous entity area known as "motivation" and aptitude assessment. Research on a small portion of any of these could fill the experimental lifetime of most researchers.

The terminal and main point of this paper concerns the role of statistics in these and all other research areas in the behavioral sciences. Statistics and experimental design are not synonomous. They are radically different matters. Experimental design is paramount, propaeduetic and foremost; statistics are a second order of business and are not essential to design. Occasionally they help, but, unfortunately they sometimes hinder, mislead and distort behavioral variations. The reader is herewith implored to concentrate his efforts on design in behavioral science and leave statistics to the statisticians.

B I B L I O G R A P H Y

Brush, F.R., Bush, R.R., Jenkins, W.O., John,W.F., & Whiting, J.W.M. Stimulus generalization after extinction and punishment: an experimental study of displacement. J. Abn. soc. Psychol., 1952, 47, 663-640.

Chapple, E.D., Chamberlain, A., Esser, A.H., & Kline, N.S. The measurement of activity patterns of schizophrenic patients. Jour. Nerv. and Men. Dis., 1963, 3, Vol. 137, 258-267.

Finan, J.L. Strength of conditioning in rats under varying degrees of hunger. J. comp. Psychol., 1940, 29, 119-134.

Jenkins, W.O. Behavior and the contiguity principle. Unpub. ms., 1955.

Jenkins, W.O. The criterion: A consideration of certain criterial matters in behavioral research. Unpub. ms., Center for Urban Educ., 1966.

Jenkins, W.O., & Mosteller, F.C. Experiments in the communication of fragmentary written messages. Unpub. ms., Univ. Tenn. 1954.

Jenkins, W.O., Pascal, G.R., & Walker, R.W.,Jr. Deprivation and generalization. J. exp. Psychol., 1958, 56, 274-277.

Locke, E.A. What's in a name? Amer. Psychologist, 1961, 16, p. 607.

Lord, D.A., & Taylor, R.E. Miller's displacement phenomenon as a simple case of generalization decrement. Unpub. ms., M.A. Thesis, Univ. Tenn. 1956.

Miller, N.E. Theory and experiment and relating psychoanalytic displacement to stimulus - response generalization. J. abnorm. (soc.) Psychol., 1948, 43, 155-178.

Miller, N.E. Comments on theoretical models: illustrated by the development of a theory of conflict behavior. J. Personal., 1951, 20, 82-100.

Miller, N.E., & Kraeling, Doris. Displacement: greater generalization of approach than avoidance in a generalized approach-avoidance conflict. J. exp. Psychol., 1952, 43, 217-221.

Morris, W.E. An experimental analysis of Charles Morris' types of discourse. Ph.D. Diss., Univ. Tenn. 1953.

Mosteller, F.C. Testing and ranking. Unpub. ms., Harvard Univ., 1950.

Murray, E.J. & Berkun, M.M. Displacement as a function of conflict. J. abnorm. soc. Psychol., 1959, 51, 47-56.

Newton, K.R. Visual recognition thresholds and learning. Perceptual and Motor Skills, 1953, 6, 81-87.

Pascal, G.R., & Jenkins, W.O. Systematic observation of gross human behavior. New York: Grune and Shatton; 1961.

Peters, C.C. & VanVoorhis, W. Statistical procedures and their mathematical bases. New York: McGraw Hill; 1940.

Rickard, H.C. The relative contribution of partial reinforcement schedule and cue change to extinction behavior. Ph.D. Diss., Univ. Tenn. 1959.

Rowe, J.M. The contiguity principle and the Skaggs-Robinson hypothesis. Unpub. Ph.D. Diss., Univ. Tenn., 1955.

Ryan, T.A. Multiple comparisons in psychological research. Psychol. Bull., 1959, 56, 26-47.

Sargent, F. An application of statistics. Sci., 1955, 121, 402.

Sheffield, F.D. "Spread of effect" without reward or learning. J. exp. Psychol., 1949, 39, 575-579.

Sheffield, F.D., & Jenkins, W.O. Level of repetition in the 'spread of effect'. J. exp. Psychol., 1952, 44, 101-107.

Sheffield, Virginia F. Resistance to extinction as a function of the distribution of extinction trials. J. exp. Psychol., 1950, 40, 305-313.

Stevenson, H.W., & Odom, R.D. Visual reinforcement with children. J. exp. child Psychol., 1964, 1, 248-255.

Weintraub, D.J., & Eisenberg, M. The shape of things to come: Miss America versus the golden rectangle. Amer. Psychol., 1966, 3 Vol. 21, 246-255.